

「電子公文書等の長期保存フォーマットを含む 長期保存に関する調査検討」の状況について

令和6年1月

独立行政法人 国立公文書館

1. 調査検討内容の振り返り

※第102回公文書管理委員会(R5.7.24)より

(1) 検討の方向性

- すべての移管文書を一律に一定のフォーマットに変換する必要はない
- 長期保存にあたってのリスクが低いフォーマットである「標準的フォーマット」(仮称)については、そのまま移管
- 行政機関等には移管文書は標準的フォーマットで作成、保存するよう求める
- 長期保存に当たってのリスクが高いフォーマットについてはフォーマットを変換して移管

(2) リスク評価の考え方

米豪の国立公文書等におけるフォーマットのリスク評価の考え方を踏まえ整理

リスク評価の考え方

- | |
|----------------------------|
| A 仕様が公開されていること |
| B 世界中で広く使用されていること |
| C 可逆圧縮等が可能なこと |
| D メタデータを付与できること |
| E 技術要素が独立していること |
| F 特許やライセンス等の制約がないこと |
| G プロテクションの制約がないこと |

(3) 標準的フォーマットの候補

米英豪の国立公文書館において長期保存上のリスクが低いとされているフォーマットを抽出し、専門家等に確認をした上で「標準的フォーマット候補」を整理

カテゴリ	標準的フォーマット候補
文書	PDF/A-1、PDF/A-2 PDF ver. 1.7以降 Microsoft Word 2007以降 (OOXML) OpenDocument Text ver. 1.2
表計算	Microsoft Excel 2007以降 (OOXML) OpenDocument Spreadsheet ver. 1.2
プレゼンテーション	Microsoft PowerPoint 2007以降 (OOXML) OpenDocument Presentation ver. 1.2
画像	JPEG 2000 TIFF 6 (Tagged Image File Format) PNG 1.2 (Portable Network Graphics) GIF (Graphics Interchange Format) JPEG (Joint Photographic Experts Group)
音声	WAVE (Waveform Audio File Format) MP3 (MPEG Layer III Audio Encoding)
動画	MPEG-4 MPEG-2 (MPEG2-PS, MPEG2-TS)

【令和5年度の実施内容】

- ① リスク評価方法案の策定
- ② 「標準的フォーマット」リスト案の策定
- ③ 「標準的フォーマット」に変換する技術等の実装に向けての検討
- ④ 行政機関等と国立公文書館との役割分担の検討

令和4年度

令和5年度

2. 令和5年度の調査検討状況

(1) リスク評価方法案の検討

- 「リスク評価の考え方」ごとの評価基準、評価点を設定
- 様々なフォーマットのリスク評価を行う「リスク評価基準表」を作成

リスク評価の考え方		評価基準	評価点		備考
A	仕様が公開されていること	仕様が公開されており、仕様を維持するための広く認められた国際的なコミュニティ活動の実態がある（例：ISO、IEC、ITU、W3C）	国際標準	2	< 3段階評価 > 評価基準に基づき、相対的なレベル設定が可能な場合は、3段階（2点、1点、0点）で評価
		仕様が公開されている	仕様公開	1	
		仕様が公開されていない、又は公開されていても広く認知されていない	非公開	0	
B	世界中で広く使用されていること	ツールが相当数あり、世界中で広く使用されている	世界中で使用	2	
		ツールが少ないが、国内もしくは一部コミュニティで一定程度使用されている	限られた使用	1	
		ツールが少なく、あまり使用されていない。又はツールのサポートが終了しているか、終了が予告されている	使われていない	0	
C	可逆圧縮等が可能なこと	無圧縮を含め圧縮品質を選択できる	圧縮可能 (圧縮品質指定可)	2	
		圧縮（可逆圧縮又は非可逆圧縮）ができる	圧縮可能	1	
		圧縮ができない、又は不明	圧縮不可、又は不明	0	
D	メタデータを付与できること	任意のメタデータを付与でき、メタデータ項目が標準に準拠している	標準準拠	2	
		任意のメタデータを付与できる	付与できる	1	
		任意のメタデータを付与できない、又は不明	付与できない	0	
E	技術要素が独立していること	特定の機器、OS等に依存しない	独立	2	< 2段階評価 > 評価基準に基づき、条件の有無により評価するものは2段階（2点、0点）で評価
		特定の機器、OS等に依存する、又は不明	依存	0	
F	特許やライセンス等の制約がないこと	特許等が存在しないか、使用許諾を得ることなく無償で 사용할 ことができる	制約なし	2	
		特許等の主張があり使用許諾が必要	制約あり	0	
G	プロテクションの制約がないこと	【評価対象外】運用リスクであるので、個別のフォーマットでは評価しない。 (長期保存上、必要に応じ、暗号化等のプロテクションを解除して保存・利用可能なことが必要)		-	

(2) 「標準的フォーマットリスト案」の検討

- リスク評価の考え方を踏まえ、「標準的フォーマットの候補」の技術情報を再整理
(各フォーマットのバージョンごとの技術的特性を確認するなどし、再整理)
- フォーマットごとに、「リスク評価基準表」に基づき評価を実施

カテゴリ	フォーマット	拡張子	評価点合計
文書	一太郎 (4~12/2004~2008)	jtd/jfw/jbw/ jaw/jsw	3
	Word 97-2003	doc	6
	OASYS	-	0
表計算	Excel 97-2003	xls	6
プレゼン	PowerPoint 97-2003	ppt	6

リスクが低いフォーマット	拡張子	カテゴリ	評価点合計	評価コメント	標準的フォーマット案
PDF/A-1	pdf	文書・ プレゼン テーション	10	ISO19005-1. PDF1.4をベースに見読性の維持に適さない機能を制限。フォント埋込必須	○
PDF/A-2			11	ISO19005-2. PDF1.7をベースに見読性の維持に適さない機能を制限。フォント埋込必須。長期保存にあたり安定しているフォーマット	○
PDF/A-3			11	ISO19005-3. PDF/A-2に加えて、多様なファイル (PDF/Aへの変換前のファイルなど) の埋込みが可能となったが、埋込みファイルの見読性やセキュリティなどに考慮が必要	
PDF/A-4			9	ISO19005-4. PDF2.0をベースに見読性の維持に適さない機能を制限している。普及は限定的	
PDF1.7			11	ISO32000-1. 広く普及している。一般的なソフトで作成可能。フォント埋込が必須ではない	○
PDF2.0			10	ISO32000-2. 最新のPDFフォーマットであるが、普及は限定的	
Word 2007以降 (OOXML)	docx	文書	11	ISO-IEC29500. 国際標準化されており、一般的なオフィスソフトで作成可能 (※1)	○
OpenDocument Text	odt		10	ISO-IEC26300. 無償で提供されているソフトで作成可能だが普及していない (※2)	
Excel 2007以降 (OOXML)	xlsx	表計算	11	(※1) と同様	○
OpenDocument Spreadsheet	ods		10	(※2) と同様	
PowerPoint 2007以降 (OOXML)	pptx	プレゼン テーション	11	(※1) と同様	○
OpenDocument Presentation	odp		10	(※2) と同様	
JPEG2000	jp2	画像	12	ISO-IEC15444-1. 可逆圧縮で圧縮品質が選択可能。長期保存の目的で広く利用	○
PNG 1.2	png		10	ISO-IEC15948. 可逆圧縮に対応。広く普及	○
JPEG	jpg		12	ISO-IEC10918. 非可逆圧縮だが圧縮品質が選択可能で、広く普及	○
TIFF 6	tif / tiff		11	仕様は公開されているが、国際標準ではない。可逆圧縮で圧縮品質が選択可能。	
GIF	gif		9	仕様は公開されているが、国際標準ではない。可逆圧縮に対応。解像度は低い、データサイズが小さくなるため普及している	
WAVE	wav	音声 動画	10	非圧縮方式であり、幅広く利用されているが、国際標準ではない。	
MP3	mp3		12	ISO-IEC11172-3. 非可逆圧縮だが圧縮品質が選択可能で、高音質。様々なデバイスで再生可能のため広く普及	○
MPEG-4	mp4		12	ISO-IEC14496-14. 非可逆圧縮だが、圧縮品質が選択可能で、高音質。広く普及	○
MPEG-2	mpg /mpeg		12	ISO-IEC13818-1. 非可逆圧縮だが、圧縮品質が選択可能で、高音質。広く普及	○

① システムへの実装に向けた文書作成時の確認

- 一般的なソフトウェアによる「標準的フォーマット」での文書作成等の点検、留意事項を把握
Word (OOXML) 等は、一般的な業務環境で作成、保存が可能。

② フォーマット判定機能

- 「標準的フォーマット」等は、英国国立公文書館が公開しているフォーマット判定ツール (DROID) をはじめ複数のツールで判定できることを確認 (「標準的フォーマット」以外を自動検出することも可能)
- 幅広いフォーマットを判定可能で、実装コストも不要な「DROID」が、フォーマット自動判定ツールとして適当

③ フォーマット変換・検証機能

- 各省庁で「標準的フォーマット」以外を保有する場合、見読性を確保するため「標準的フォーマット」へ変換しておくことが望ましい。そのため、「標準的フォーマット」以外を、「標準的フォーマット」への自動変換・自動検証する試験を実施
 - ・ 試験環境では「標準的フォーマット」への個別変換の自動化は可能 (※メーカーで推奨・保証するものではない。)
 - ・ OASYSは変換不可。複数シートの一太郎は手動変換のみ可
 - ・ 変換結果の検証機能については、
変換前後の文字 (テキストの比較)、表示内容 (レイアウト) の比較ツールにより一定程度は可能
 - ➡ ✓ 変換・検証の精度や信頼性については更なる研究が必要
 - ✓ システム化に向けて、必要となる処理能力や実装コストの確認が必要

④技術等の確認結果からの論点

フォーマット判定について

- 論点1 新規作成する電子公文書については、記録領域に保存時に「標準的フォーマット」かどうかをチェックする機能を設けてはどうか。
- 論点2 行政機関等での見読性確保の観点から、既に保存している「標準的フォーマット」以外の電子公文書について把握する機能を設けてはどうか。
(※不要なファイル：一時保存ファイルやジャンクファイルの点検含む)
- 論点3 見読性の確認機能（再生確認）を設けてはどうか。

【不要なファイルの例】
～\$公文書管理.docx
○○○.sys
_公文書管理.xlsx
公文書管理. \$\$\$

フォーマット変換・検証について

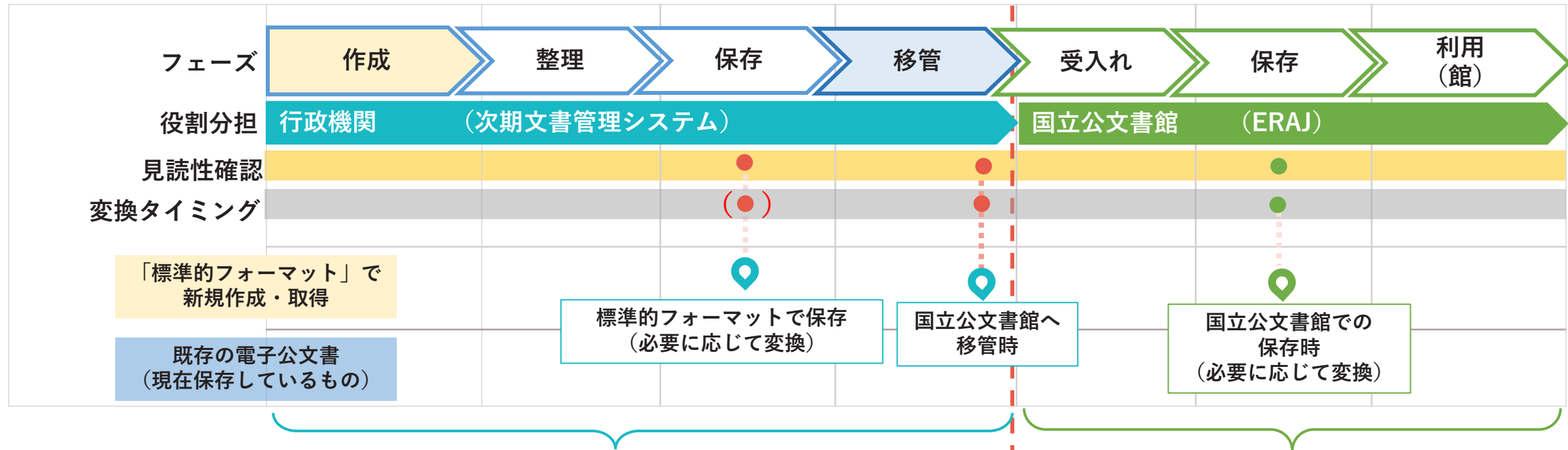
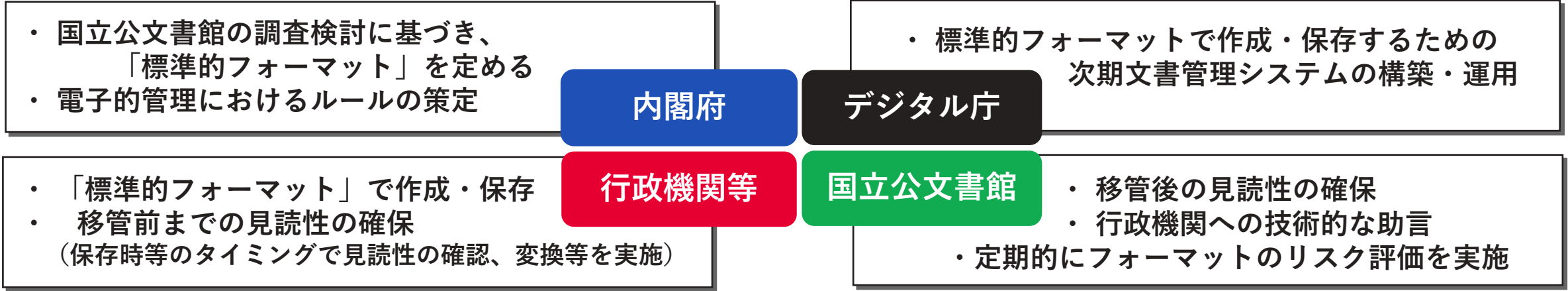
- 論点4 自動変換・検証機能についてシステム化及び技術の確立にむけた調査研究が必要。
- 論点5 今後自動変換を実現した場合も変換前後において文書の本質的な内容が失われていないかの確認する必要があると考えられる。

「標準的フォーマット」以外のフォーマットへの対応

- 論点6 変換できないフォーマットの保存・利用のため、引き続き継続的に検討が必要（仮想的な再生環境等）。

(4) 行政機関等と国立公文書館との役割分担の検討

○ 基本的な考え方



行政機関にて見読性を確保※

国立公文書館にて見読性を確保

※保存期間が長期にわたる場合には、保存期間中にも定期確認し必要に応じて変換をすることを検討

3. 今後について

(1) 令和5年度

- 国の次期文書管理システムで必要となる機能の実装に向けて、要件の具体化

(2) 令和6年度

- 「標準的フォーマット」で移管、保存したものを利用に供する方法の検討
 - 利用制限情報のマスキング方法、電子的に紛れ込んでしまう情報の確認方法等
- 「標準的フォーマット」による保存等の留意事項に関する技術的検証
 - 「標準的フォーマット」であっても、特定の機能（関数、外字、異体字、電子署名等）を使用していた場合の長期保存の安定性について調査
- 「標準的フォーマット」以外のフォーマットの保存及び利用に向けた調査研究の継続
 - 「標準的フォーマット」以外のフォーマットの自動変換及び自動検証、仮想的な再生環境に関する調査研究